DOCUMENT RESUME

ED 246 664 FL 014 448

AUTHOR Cochran, Judith; Mills, Carl
TITLE ESL Teacher Effectiveness as Measured by Student Performance on the English Language Proficiency Test.
PUB DATE [83]
NOTE 16p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Academic Achievement; Classroom Techniques; *English (Second Language); Foreign Countries; *Language Tests; Measurement Techniques; *Second Language Instruction; *Teacher Effectiveness; *Teacher Evaluation; *Teaching Skills
IDENTIFIERS American University (Egypt)

ABSTRACT

This study measured the usefulness of a competency-based language test to identify and measure the specific teaching competencies of English language teachers at the American University in Cairo. Since a major concern was to employ a methodology easily used in an ongoing program, an established program component, the English Language Proficiency Test, was used to measure student achievement, while the Instrument for Observing Teacher Activities, which could be used reliably among all classes, was used to measure teacher competency for correlational purposes. Thirty-one English language teachers in the Department of Public Service and their 333 students were the research sample. Statistical analysis of the correlations between the student and teacher measures revealed no significant relationship between nine indicators of teacher competency and high or low student achievement. Explanations may include too small a sample, bias introduced by selecting only high and low student averages for the sample, use of teacher competencies that are not relevant, and use of an inappropriate measure of student performance. It is also suggested that the best teachers may simply not "teach to test." Further research into the factors behind student success on proficiency tests as they relate to teacher competency is recommended. (MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Judith Cochran

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

ED246664

FL014448

ESL TEACHER EFFECTIVENESS AS MEASURED BY STUDENT PERFORMANCE ON THE ENGLISH LANGUAGE PROFICIENCY TEST

Judith Cochran and Carl Mills

English language proficiency tests such as the TOEFL, Michigan and ALIGUE are given to thousands of ESL students every month throughout the world. Students' scores on these tests determine college entrance, scholarship eligibility and to a large extent measure ESL programs' success. The assumption is that successful programs also have competent teachers whose instructional processes enable their students to succeed on these proficiency exams. But the teaching behaviors these so called successful teachers are exhibiting remain largely unexamined for a number of reasons; few studies have examined observed teacher competencies as related to student achievement on proficiency tests, and few language programs provide a large enough sample to begin to gather baseline data on either student preformance or teacher competency. This two-year study proposed to do just that; to examine the specific teaching competencies demonstrated by successful teachers as determined by student achievement on a standardized English proficiency test.

The impetus for this study began in 1980 when all elementary teachers in the seven educational zones in Cairo were observed using a competency base. observation instrument developed for the Ministry of Education.

Results of this pilot study indicated that the instrument did in fact reveal different instructional competencies between successful math and language teachers as determined by student achievement on national math and language proficiency tests. (Cochran, Ebied, Mofti, 1981, p. 1-15). This study provided the basis for assuming that a profile of competencies for English language teachers could be identified through the use of a competency based instrument.

The studies which have examined teacher competencies as determined by student achievement have focused upon both the usefulness and difficulty in conducting such studies. They have viewed tests as a means for revealing variations in the effectiveness of different curricula, motivating teachers and students to succeed on a specific goal, recognizing and rewarding success in learning and focusing on schooling as a purposeful educational activity that can be assessed and documented systematically. (Ebel, R. 1977, p. 1-51).

A study conducted by the California Commission for Teacher Preparation and Licensing developed an assessment system for measuring student and teachers' behavior in the classroom in a two-phase study used for beginning teacher evaluation. Actual classroom teaching was observed; pupil performance was measured in reading and mathematics on 41 second and 54 third grade experienced teachers. The limited historical data available from such evaluation classrooms, made it inappropriate to formulate any hypothesis from this study which involved the collection data on students in the classroom of the total 95 teachers observed over a two year period. (Wheeler, P. 1976, p. 25).

It has been suggested that traditional methods of evaluating teachers are not adequate. For example, evaluators' judgements are usually too subjective to provide valid data while standardized tests are often too insensitive to student performance on the specific goals and objectives of a given program. (Atkin, M., 1972, p. 11). Student performance on relevant measures should be used as the primary basis for teacher evaluation systems.

However, the difficulties in measuring the interrelationship between student performance and teaching behavior involves the problems of in-

strumentation, methodology, and evaluation. There is the obvious problem of the unknown predictive validity of standardized tests from specific teaching units, the problem of building multivariate outcome measures and the problems of measurement of appropriateness of teacher behavior, the lack of experience in choosing an appropriate unit of analysis for describing teacher behavior and the lack of stability of many teacher behaviors, (Berliner, D., 1975, p. 34). These problems were encountered by Baker and Popham's' performance test of teaching effectiveness conducted in 1966. Scores on student achievement measures on two industrial education topics were assessed for reliability and intercorrelated with measures of grade point averages, student's interest in subject matter and Wonderlic Personnel Test scores. The objective was to detect variables that could potentially be used to control for student differences in preformance in assessing teacher effectiveness. Kuder-Richardson reliability coefficients of .44 and .78 were found for the achievement tests. Test scores correlated .68 with grade point average.

The major concern of this ESL study was to conduct research with methodology easily used in an on-going program. This constraint limited the assessment of teaching behaviors to a single instrument which could be used with reliability among all classes.Furthermore, since one proficiency test, English Language Proficiency Test (EPET) was administered for the end of stage promotion in all classes, it was an established component of the evaluation process. While two measurement instruments might have seemed inadequate for assessment, they both had historical documentation within the program and were simple enough for usage in other programs should results indicate a research direction.

And finally, since no other studies on ESL instruction had investigated what teachers do that significantly affects successful student performance on proficiency tests, a pilot study begun in this fashion seemed worthy of investigation.

POPULATION

The total population consisted of ninety seven English language instructors teaching 4,142 adult Egyptian students in the Department of Public Service at the American University in Cairo. From these ninety-seven instructors, those whose classes had the highest and lowest averages on the English Language Proficiency Test became the teacher research population (n=31). Each class had an enrollment of approximately 23 students for a total of 333 students in the research population.

INSTRUMENTATION

The Insrument for Observing Teacher Activities (I.O.T.A.) was used to measure teacher competency. The IOTA has a reliability of .85 for 27 scales. The nine scales selected from the instrument measured Variety of Teaching Activities, Learning Difficulties, Classroom Control, Use of Materials for Instruction, Opportunity for Participation, Response to Student Opinion, Development of Student Initiative, Social Climate, and Subject Matter Preparation. The nine scales were selected as the most appropriate scales for measuring teaching activities in the Egyptian setting. Six Egyptian and American educators made the scale selection based upon experience in Egyptial education and 1978 and 1979 research conducted on D.P.S. teacher behaviors. (Lawrence, M., 1980 p. 1-30). The basic assumptions underlying the instrument construction are the measurement of teaching activities rather than the teacher characteristics data collection which results in objectivity vs. subjectivity and the definition of criterion referenced evaluation.

The D.P.S. English Language Proficiency Test was used to measure student achievement. The test components measure grammar, structure, and reading comprehension. The test has been administered to 40,000 Egyptian students between 1979-1983. Cut off points were determined by classical statistics and validated again by Rauch analysis. Items for the test were submitted by D.P.S. English language teachers and item selection was refined through repeated administration over a four year period. In one small study (n=61), the English Language Proficiency Test had an .87 correlation with the TOEFL.

PROCEDURES

From 1981-1982, English language teachers who taught intermediate English (level 33) at the D.P.S. English language program at American University in Cairo were observed on the same instrument to determine if a teaching competency profile could be identified. All students on level 33 were tested at the end of 36 hours of instruction on the English Language Proficiency Test. The profile of teaching competencies which emerged was tentative as it involved only the 5 highest and 5 lowest class instructors. The study was expanded in 1982-1983 to include teachers of all classes taking the English Language Proficiency Test. Five differenct courses were examined: beginning (11, 12, 13) and intermediate (41 and 43). Class scores on the EPET were averaged with the research population consisting of the 31 high and low classes taught by 31 different instructors. To determine the consistency of the teaching profile, each teacher had been observed twice by trained teacher observers who had an inter-rater reliability of .85.

All class content came from a single English textbook series, Intercom. Each teacher had a course syllabus of behavioral course objectives which was followed carefully be teachers to enable students to succeed on departmental achievement tests which were also administered at the end of the course. The achievement tests were developed from the text while the proficiency test was secured in the department and was administered and scored in seperate arrangements. This situation closely parallels the administration of other proficiency tests such as the TOEFL, ALIGUE and Michigan Exam. The scores on nine scales of teacher behavior became the independent variable and student achievement became the dependent variable in assessing teaching competencies.

RESULTS

Statistical analyses of data on observed teacher competencies for 9 criteria and class averages revealed that there is no significant correlation between the 9 areas of teacher competency and high or low average student performance on standardized tests.

Specifically, for 8 classes at Level 11, 8 classes at Level 12, 6 classes at Level 13, 6 classes at Level 41, and 3 classes at Level 43, a number of statistical analyses were performed, none of which showed any significant association between student performance and teacher evaluation.

Since teacher competency is a continous variable, presumed to be measured on an interval scale, an attempt was made to analyze average teacher competency as a dependent variable with Level (11, 12, 13, 41, 43) and dichotomized student class averages ("high" or "low"), referred to here as "HILO" as independent variables.

First, a two-way Analysis of Variance (ANOVA) was performed using the Statistical Package for the Social Sciences (SPSS). Results of the two way ANOVA are given in Table 1.

TABLE 1 2 WAY ANOVA
COMPETENCY BY AVERAGE LEVEL

Source of Variation	Sum of Squares	DF	Mean Square	F	Signif of F
Main Effects	82.088	5	16.418	0.459	0.802
HILO	1.200	1	1.200	0.034	0.856
Level	81.551	4	20.388	0.570	0.687
2-way Interactions	148.092	4	37.023	1.035	0.412
HILO Level	148.092	4	37.023	1.035	0.412
Explained	230.180	9	25.576	0.715	0.690
Residual	750.914	21	35.758		
Total	981.094	30	32.703		

These results indicate that one cannot find a significant difference in teacher competency by either level or class average (HILO). Teacher competencies were the same from level to level and for classes with high or low averages. Furthermore, there was no significant interaction between level and average. A Oneway ANOVA by levels confirmed that there was no significant difference in average teacher competency from level to level.

TABLE 2 COMPTENCY BY LEVELS

Source of Variation	F	Sum of Squares	Mean Squares	F	F Prob
Between Groups	4	80.8951	20.2238	0.584	0.6769
Within Groups	26	900.2079	34.6234		
Total	30	981.1028			

A two-tailed t-test confirmed that there was no significant difference in average teacher competency between classes with high averages and classes with low averages (t=.13; p=.901).

Since average teacher competency is a continous interval variable and class average (HILO) represents a dichotomization of an underlying continous variable, correlations between average teacher competency and class average were analyzed.

The Pearson Product Moment Correlation Coefficient r (SPSS) between average teacher competency and class average was -.0234, which was not significant (p=.450).

Because class average had been dichotomized into "high" and "low" averages, the Pearson r was not, strictly speaking, the proper measure of correlation, though it proved to be a remarkably robust statistic in this case. Accordingly, a Point Biserial Correlation Coefficient was calculated by hand between average teacher competency and class average. In this case rpbi=-.02344, which differed from r by only .00004. Neither r nor rpbi was significant.

Up to now, we have been dealing with a teacher competency rating calculated by averaging ratings for 9 seperate competency measures that were actually rated by classroom observers. Analyses of the relations between specific teacher competencies and class averages (HILO) proved equally fruitless for establishing significant correlations.

The 9 measures of teacher competency (Variety in Teaching Activities, Learning Difficulties, Classroom Control, Use of Materials for Instruction, Opportunity for Participation, Response to Student Opinions, Development of Student Initiative, Social Climate, and Subject Matter Preparation were first analyzed for intercorreleations among themselves. Results are given in Table 3.

Table 3 Pearson Correlation Coefficients for Competencies

	C1	C2	C3	C4	C5	C6	C7	C8	C9
C1.	1.0000	0.2480	0.3995	0.5853	0.5674	0.6390	0.7562	0.6534	0.6925
	p=****	p=.089	p=.013	p=.000	p=.000	p=.000	p=.000	p=.000	p=.000
C2		1.0000	0.3957	0.1826	0.1496	0.3220	0.3085	0.2332	0.3096
		p=****	p=.014	p=.163	p=.217	p=.039	p=.046	p=.103	p=.045
C3.			1.0000	0.3136	0.3161	0.3598	0.4173	0.5004	0.3260
			p=****	p=.043	p=.042	p=.023	p=.010	p=.002	p=.037
C4.				1.0000	0.3486	0.4197	0.6853	0.4734	0.5108
				p=****	p=.027	p=.009	p=.000	p=.004	p=.002
C5.					1.0000	0.6304	0.5977	0.6259	0.4966
					p=****	p=.000	p=.000	p=.000	p=.002
C6.						1.0000	0.8150	0.7034	0.7368
						p=****	p=.000	p=.000	p=.000
C7.							1.0000	0.7277	0.7542
							p=****	p=.000	p=.000
C8.								1.0000	0.7403
								p=****	p=.000
C9.									1.0000
									p=****

As is readily apparent, the individual measures of teacher competency are highly correlated with one another. Of the 36 relevant intercorrelations, 16 have $p < .001$, 6 have $p < .01$, and 10 have $p < .05$.

Once again, though the Pearson r is not the most appropriate measure of correlation between continous variables like the 9 teacher competencies and a dichotomous variable like class average, the ease with which r can be calculated using SPSS and the closeness

of fit between r and the more appropriate Point Biserial Correlation Coefficient rpbi led us to calculate r for all teacher competencies and class average. These correlations are given in Table 4.

Table 4. Pearson Correlation Coefficient: Comp and Average

	C1	C2	C3	C4	C5	C6	C7	C8	C9
Av.	.0126	-.0775	-.0661	.0083	-.0535	.0711	.0489	.0054	-.0083
p=	.47-	p=.339	p=.362	p'.482	p=.388	p=.352	p=.397	p=.488	p=.482

As Table 4 shows, there was no significant correlation between any of the 9 measures of teacher competency and high or low class average.

Finally, Point Biserial Correlations were calculated by hand for each of the 9 teacher competencies and class average. These correlations are given in Table 5.

Table 5. Point Biserial Correlations: Comp with Average

C1	C2	C3	C4	C5	C6	C7	C8	C9
Av.0123	-.-763	-06501	.0081	-.0527	.750	.1110	.0091	-.0081

(None of these values for rpbi is significant.)

Two things need to be said of Table 5.

First, the Point Biserial Correlation Coefficients agree quite well with the corresponding Pearson r's in Table 4. In our opinion, this simpl- illustrates the robustness of r. Second, none of the Point Biserial Correlation Coefficients was significant, demonstrating once again that there was no significant correlation between the 9 measures of teacher competency and high or low average student performance.

On the basis of these statistical analyses we conclude that the Null Hypothesis-that teacher competency ratings have no effect on student performance ratings-cannot be rejected.

It remains for us to interpret these results. We offer four possible explanations, which are not mutually exclusive, for our findings.

First, we analyzed data from 31 classes taught by 31 teachers. It may simply be the case that our sample was too small. A larger sample of classes, teachers, and levels might yield different results. Obviously, this study should be regarded as a pilot study, and more research is needed.

Second, we analyzed data from classes with the highest and lowest student averages. And these data on averages were dichotomized into "high" and "low" categories. This dichotomization resulted in a considerable loss of information. It is possible, perhaps likely, that utilization of the full range of available actual class averages (avoiding dichotomization of the data on class average) might result in different correlation coefficients. A new study needs to be conducted to minimize loss of information and to procure accurate correlation coefficients.

Third, the measures of teacher competencies used may be wrong or, more precisely, irrelevant. In fact, at present, we are not inclined to accept this conclusion: the nine factors we studied simply do not measure effective teaching very accurately. Another criteria on which teachers classroom performance can be evaluated may need determined.

Finally, it may be that the measure of student performance (class average) is inappropriate. Put bluntly, perhaps we used the right criteria to evaluate the teachers' techniques but the wrong criteria to evaluate the effectiveness of these teaching techniques.

Perhaps our best teachers simply do not "teach to the test," and perhaps we need to use other measures of student performance in evaluating our teachers. This is, of course, a common problem with standardized tests which needs to be considered in student and teacher evaluation.

For now, we conclude that even the negative findings reported here have value in indicating directions for future research and in pointing out some of the pitfalls researchers can encounter when evaluating teaching effectiveness. One might hypothesize that successful teachers teach to the test. And if structures on proficiency tests are unknown to teachers, as on the EPET, success cannot be predicted in any objective manner. Furthermore, if textbooks do not provide instruction relevant to proficiency tests, a great deal of effort is wasted in the classroom if success is measured on student proficiency scores. The students' success on proficiency tests could be as our study reveals, merely a result of chance, not a result of teaching competencies.

BIBLIOGRAPHY

1. Baker, Eva, Popham, Janus, (1966) "A Performance Test of Teaching Effectiveness." Paper presented at 1966 American Educational Research Association Meeting, Chicago, February 17-19, 1966.

2. Berliner, David (1975), "Impediments to the Study of Teacher Effectiveness," Far West Laboratory for Educational Research and Development, San Francisco, CA.

3. Ebeid, William, Cochran, J., El-Mofti, M.,"Teaching Compentencies Which Affect Primary School Achievement," Journal of Education, Ain Shans University, 1981.

4. Ebel, Robert L. (1977), "The Uses of Standardized Testing," Fastback Series #93, Phi Delta Kappa Educational Foundation, Bloomington, Indiana.

5. Wheeler, Patricia, Elias, Patricia (1976), Beginning Teacher Evaluation Study: Phase II, 1973-74, Final Report Vol. 3. Historical Document Study, "California State Commission for Teacher Preparation and Licensing, Sacramento; Educational Testing Service, Princeton, N. J.